Available online at www.sciencedirect.com
SCIENCE DIRECT®
Metabolic Engineering 7 (2005) 27–37
METABOLIC ENGINEERING
www.elsevier.com/locate/ymben

Semi-synthetic DNA shuffling of *aveC* leads to improved industrial scale production of doramectin by *Streptomyces avermitilis*

Kim Stutzman-Engwall[a,*], Steve Conlon[a], Ronald Fedechko[a], Hamish McArthur[a], Katja Pekrun[b], Yan Chen[b], Stephane Jenne[c], Charlene La[c], Na Trinh[c], Seran Kim[c], Ying-Xin Zhang[c], Richard Fox[c], Claes Gustafsson[b], Anke Krebber[c]

[a]*Pfizer Global Research & Development, Gorton Laboratories, Eastern Pt. Rd. MS 4123, Groton, CT 06340, USA*
[b]*Maxygen, 515 Galveston Drive, Redwood City, CA 94063, USA*
[c]*Codexis, 200 Penobscot Drive, Redwood City, CA 94063, USA*

Received 6 April 2004; accepted 7 July 2004
Available online 22 January 2005

Abstract

The avermectin analog doramectin (CHC-B1), sold commercially as Dectomax™, is biosynthesized by *Streptomyces avermitilis*. *aveC*, a gene encoding an unknown mechanistic function, plays an essential role in the production of doramectin (avermectin CHC-B1), modulating the production ratio of CHC-B1 to other avermectins, most notably the undesirable analog CHC-B2. To improve the production ratio for doramectin, the *aveC* gene was subjected to iterative rounds of semi-synthetic DNA shuffling. Libraries of shuffled *aveC* gene variants were transformed into *S. avermitilis*, screened using a miniaturized 96-well growth and production format, and analyzed by high throughput mass spectrometry to determine CHC-B2:CHC-B1 ratios. Several improved *aveC* variants were identified; the best shuffled gene encoded 10 amino acid mutations, and conferred a final CHC-B2:CHC-B1 ratio of 0.07:1, a 23-fold improvement over the starting gene (*aveC* wild type). Chromosomal insertion of an improved *aveC* shuffled gene into a high titer *S. avermitilis* strain yielded an improved doramectin production strain. This strain is under development to be used commercially, and is expected to provide considerable cost savings in large-scale manufacture, as well as significantly reducing by-product levels of CHC-B2 requiring disposal.

Keywords: *Streptomyces avermitilis*; *aveC*; Semi-synthetic DNA shuffling

1. Introduction

Polyketides are a large and structurally diverse group of natural products with activities ranging from antibacterial (erythromycin), anticancer (daunorubicin), antifungal (amphotericin), cholesterol-lowering (mevastatin), immunosuppressant (rapamycin) to anthelmintic (avermectin). The highly complex polyketide carbon chain backbone is produced by the sequential activity of the multi-enzyme and often multi-modular polyketide synthase. Variations in the number of extensions, choice of starter unit and extender units, stereochemistry, reduction of the polyketide chain, and subsequent modification of the polyketide backbone by methyl-, sugar-, and hydroxyl groups, all contribute to the structural diversity of these types of compounds (Katz and Donadio, 1993). As a class, polyketides are one of the richest sources of pharmaceuticals today. Consequently, considerable current research is focused on manipulation and design of polyketide synthetic pathways (Chartrain et al., 2000).

Avermectins are 16-membered macrocyclic lactones produced by *Streptomyces avermitilis* (Burg et al., 1979). The natural and semi-synthetic derivates of avermectin

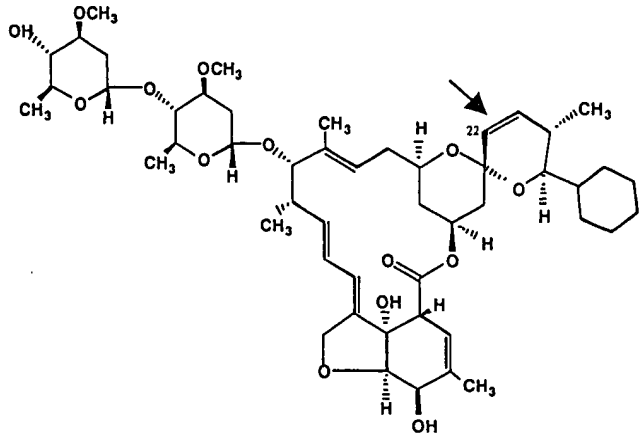
*Corresponding author.
E-mail address: stutzmanengw@groton.pfizer.com (K. Stutzman-Engwall).

1096-7176/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2004.07.003

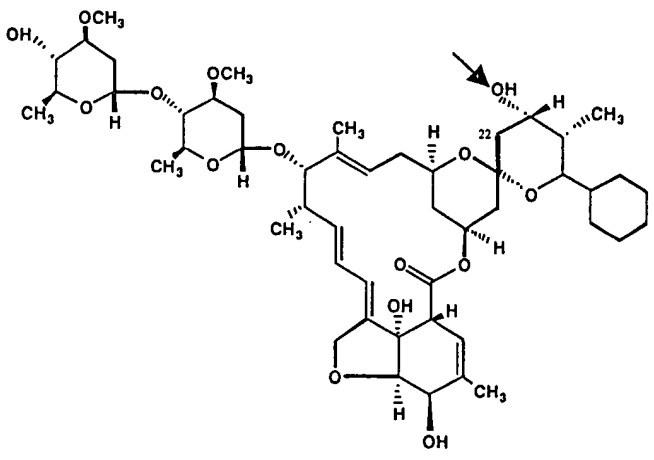
are potent anthelmintic and insecticidal compounds and are widely used as veterinary treatments for broad spectrum parasite control and in human medicine for the treatment of river blindness (onchocerciasis). Avermectin derivatives are the most widely used drugs in animal health and agriculture, with current worldwide sales exceeding 1 billion US dollars. The avermectin polyketide chain is derived from seven acetates and five propionates, together with a single 2-methylbutyric acid or isobutyric acid residue which forms the *sec*-butyl or isopropyl group attached to C25 (Cane et al., 1983; Chen et al., 1989). The polyketide chain is subsequently cyclized, glycosylated at C13 with the attachment of two oleandrose sugars and methylated at the hydroxyl group of C5. In the biosynthesis of the analogue doramectin, cyclohexanecarboxylic acid (CHC) is used in place of the natural chain initiator. The doramectin production strain is a mutant of *S. avermitilis* lacking branched-chain-2-oxo acid dehydrogenase (*bkd*) and *O*-methyltransferase activity. This strain produces only the B forms of avermectin corresponding to the supplemented fatty acid (Dutton et al., 1991; Hafner et al., 1991).

The *S. avermitilis* doramectin production strain synthesizes a mixture of two related compounds, CHC-B1 (doramectin) and CHC-B2. The CHC-B1 compound has a double bond between C22 and C23 whereas CHC-B2 contains a saturated C–C bond with a hydroxyl group at C23 (Fig. 1), and it has been proposed that the CHC-B2 form is converted to the B1 form by a dehydration event at C22,23 (Chen and Inamine, 1989). CHC-B2 is a less effective pharmaceutical agent than CHC-B1 (Goudie et al., 1993). Although characterization of the genes in the avermectin gene cluster has clarified much of the biosynthetic pathway (Ikeda et al., 1999; Ikeda and Omura, 1995; MacNeil et al., 1992), the mechanism that determines the production ratio CHC-B2:CHC-B1 remains unclear. The presumptive dehydration branch point resulting in the differentiation between the B1- and B2-type compounds is proposed to be an early event in the biosynthetic pathway; probably occurring during the PKS catalyzed condensation (Chen and Inamine, 1989). We have determined that the dehydratase function in module 2 of the PKS that would correspond to C22–23 dehydration does not have a role in determining CHC-B2:CHC-B1 ratio (Stutzman-Engwall et al., 1997). In contrast, when a portion of the *aveC* gene was randomly mutated, a single avermectin component B2a was produced (Ikeda et al., 1995). Deletion of *aveC* abolishes production of significantly detectable amounts of both avermectins B1 and B2, whereas overexpression of *aveC* has no effect on ratio in wild-type strains (Stutzman-Engwall et al., 2003). This *aveC* deletion mutant can be fully complemented *in trans*. Several unrelated mutations were identified after site-specific mutagenesis and error-prone PCR that resulted in significantly improved ratios of doramectin to CHC-B2 produced (Stutzman-Engwall et al., 2003). The majority of the observed random mutations, however, influence the production ratio adversely (in favor of CHC-B2).

We report here further improvements in production ratios obtained by evolving *aveC* by semi-synthetic DNA shuffling and screening for variants producing increasing amounts of doramectin relative to CHC-B2. This semi-synthetic shuffling format improved the distribution of recombinant genes in shuffled libraries, which reduced the number of variants that needed to be screened to identify improved variants. The best variants identified from the first round of semi-synthetic shuffling produced a CHC-B2:CHC-B1 ratio of 0.2:1, while retaining the same total avermectin (CHC-B2 + CHC-B1) yield as the wild-type gene. Subsequent rounds of



CHC-B1 (Doramectin)



CHC-B2

Fig. 1. Chemical structure of the CHC-B1 (Doramectin) and CHC-B2 avermectin analogues. Arrows mark the enoyl carbons at C22-23 in the CHC-B1 analogue and the hydroxyl group at C23 in the CHC-B2 analogue.

semi-synthetic gene shuffling yielded further improvements, resulting in an *aveC* variant with a CHC-B2:CHC-B1 ratio of 0.07:1. The best *aveC* gene was subsequently introduced into the chromosome of the production strain; this new production strain achieved a maximal ratio improvement.

2. Materials and methods

2.1. General methods

Escherichia coli DH5α was used for routine subcloning employing standard growth media and conditions (Sambrook et al., 1989), or as described in suppliers' protocols. *E. coli* transformants were selected with 100 μg/ml ampicillin. *S. avermitilis aveC* mutant SE180-11 (Stutzman-Engwall et al., 2003) containing an insertionally inactivated *aveC* gene, was used as host for complementation experiments. Plasmid pWHM3 (Vara et al., 1989) was used for routine subcloning for transformation into *S. avermitilis*. Plasmids were isolated from *S. lividans* or *E. coli* DM1 (BRL) or *E. coli* SCS110 (Stratagene) prior to transformation into *S. avermitilis*. *S. avermitilis* routine and high throughput (HTP) culture conditions were similar to those described previously (Stutzman-Engwall et al., 2003). *S. avermitilis* was grown for fermentation, and assays performed as previously described (Pacey et al., 2000; Stutzman-Engwall et al., 2001). DNA was sequenced using a *Taq* Dye Deoxy terminator cycle sequencing kit and an ABI 373A Automated DNA Sequencing system (Perkin Elmer, Foster City, CA). Sequence data were assembled and edited using Genetic Computer Group programs (GCG, Madison, WI, Devereux et al., 1984). PCR amplification and genomic replacement protocols were similar to those described previously (Stutzman-Engwall et al., 2003).

2.2. Plasmid and strain constructions

S. avermitilis aveC mutant SE247-11 (Stutzman-Engwall et al., 2003) was used as the production strain host to introduce gene replacement vectors containing mutated *aveC* DNA. Gene replacement vector pSE370 was constructed in a 4-way ligation containing a 0.6 kb *Sph*I/*Pst*I fragment from pSE617 (CHC-B2:CHC-B1 ratio of 0.05:1), a 1.4 kb *Pst*I/*Bam*HI fragment from pSE180, a 2.8 kb *Xba*I/*Sph*I fragment from pSE180 and *Bam*HI/*Xba*I-cleaved pWHM3. Because pSE370 was missing *aveC* mutation E238D, a second gene replacement vector pSE375 was constructed. pSE370 was used as a template to PCR a ~600 bp *Sph*I/*Nsi*I fragment using a mismatched primer (5′-CTGCTGCATGCTGGGCTCGCTGCGCTTCTTCCGCGACGACC-3′) and reverse primer (5′-GCTGGGAAACCAAGGGATCG-3′). The PCR product was digested with *Sph*I/*Nsi*I and a 3-way ligation with a 9.2 kb *Sph*I/*Hind*III fragment and a 2.5 kb *Hind*III/*Nsi*I fragment was performed. All plasmid constructions were confirmed by restriction analysis and, in some cases, DNA sequence analysis.

2.3. Semi-synthetic shuffling

Previously, mutation of the *aveC* gene yielded 7 variants, each conferring a reduced CHC-B2:CHC-B1 ratio while maintaining overall avermectin yield (Stutzman-Engwall et al., 2003). DNA sequence analysis determined that each clone encoded between two and four AA substitutions. The best first round clone (pSE290, CHC-B2:CHC-B1 ratio of 0.4:1) was shuffled using a method described by Stemmer (1994b). Individual oligonucleotides encoding AA substitutions corresponding to single mutations conferring an improved B2:B1 product ratio were added to the shuffling reaction at 5 molar excess over *aveC* template DNA. In the case of AA136, AA138, and AA139, pairs of AA substitutions were constructed in a single oligonucleotide. Each nucleotide mismatch in an oligonucleotide was flanked by 20 nucleotides of perfect identity to optimize incorporation during the shuffling reaction. Oligonucleotides were purchased from Operon technologies (Alameda, CA). Subsequent libraries were generated using best available full length *aveC* variants as the shuffling template, and oligonucleotides encoding other beneficial mutations were added to the reaction. Mutazyme (Stratagene) was added to the reaction when additional diversity from random DNA mutations was desired.

2.4. Growth of clones for HTP analysis of doramectin production

Shuffled libraries were cloned into pWHM3 and transformed into methylation negative *E. coli* INV110 cells (Invitrogen). DNA sequencing of at least 10 randomly chosen colonies verified the quality of a shuffled library. Plasmid DNA from ~10^5 pooled clones was isolated and transformed into *S. avermitilis* SE180-11 protoplasts. Individual clones were grown at 30 °C (200 μL R5 medium (Kieser et al., 2000) containing 5 μg/ml thiostrepton in deep 96-well seed plates in a humidified shaker at 250 rpm). A glass bead was included in each well for dispersion of mycelia and agitation of the culture. After 3–5 days, 20 μl of the seed culture was transferred to production plates. The remainder of the seed plate was stored as a master plate at 4 °C. Production plate media were prepared essentially as described in Pacey et al. (2000), with the addition of 1% agarose. The 96 well production plate (containing 1 ml production medium per well) was tilted during solidification of the medium to increase surface

area for growth. Inoculated production plates were incubated at 30 °C under humidity for 12–14 days.

2.5. Extraction and ESI-MS/MS screening

Ethyl acetate (1 ml) was added to each well, and the plate was shaken at room temperature for 20 min. Approximately 750 µl of the ethyl acetate phase was recovered, transferred to a 96-well plate, and allowed to evaporate over night. The residue was resuspended in 100 µl methanol containing 1 mM NaCl; 5 µl was injected directly into a triple-quadrupole mass spectrometer (Micromass Quattro LC) from the 96-well format using a TwinPal LEAP auto sampler. Products were analyzed in flow injection mode without prior liquid chromatography separation. This method allowed rapid (<25 s per sample), quantitative and precise detection of products by detecting the molecular weight of the product, as well as a characteristic fragmentation ion produced in the third quadrupole of the mass spectrometer. The MS/MS transition for CHC-B1 sodiated ion is from *m/z* 921 to *m/z* 777; for CHC-B2, the sodiated ion is from *m/z* 939 to *m/z* 795 in positive mode. Integration of the separate CHC-B1 and CHC-B2 chromatograms for each well identified clones with altered CHB-B2 over CHB-B1 ratios. Clones producing improved ratios were retested in triplicate starting from the original seed plate using the same growth, production and assay protocol as described above. Clones producing improved ratios were confirmed by fermentation in 30 ml shake flask, under conditions described previously (Stutzman-Engwall et al., 2001).

3. Results

3.1. Semi-synthetic shuffling of improved avec clones

In certain previously described formats for DNA shuffling, each improved gene is isolated and fragmented, and a mixture of fragments from each gene is recombined to create the next round library (Stemmer, 1994a, b). Recombination of variants by DNA shuffling creates novel permutations of substitutions, however, to adequately explore the potential genetic diversity of a given library, a very large number of variants from that library must be screened (Moore et al., 1997). In particular, to identify clones possessing higher numbers of substitutions conferred by distinct parental gene sequences, large libraries may either be screened extensively, or multiple iterative rounds or DNA shuffling and screening of smaller numbers of variants may be pursued.

In the present study, due in part to long growth periods required for *S. avermitilis*, the screen is both complex and time-consuming (a single round of shuffling, transformation, growth, production and screening of 5000 variants, followed by characterization of potential hits can require approximately 2–3 months). The implementation of a new, semi-synthetic shuffling (oligonucleotide-based) format (Crameri et al., 2000) allowed for the exploration of diversity afforded by all variations identified in the initial screen of first round variants in a much smaller library. The theory behind the semi-synthetic approach pursued here is similar to other techniques that can be used to accelerate the recombination of beneficial diversity (Aita et al., 2002; Vajdos et al., 2002).

Ten amino acid substitutions present in the seven first round hits generated by error prone mutagenesis corresponded with improved CHC-B2:CHC-B1 ratios (Stutzman-Engwall et al., 2003). Oligonucleotides encoding individual substitutions (A61T, F99S, G111V, L136P, G179S, V196A, E238D, P289L; plus all combinations of L136P/S138T/A139T and A139F) were designed so that all combinations of amino acid changes would be afforded with equal probability in the resulting library. Oligonucleotides were mixed and shuffled with fragmented *aveC* derived from the best first round variant, pSE290 (D48E, A89T), which produced a CHC-B2:CHC-B1 ratio of 0.4:1.

Fifteen randomly isolated clones from the library were sequenced to determine oligonucleotide incorporation and error frequency. All but one of the substitutions were incorporated in 12 clones at least once (the average number of mutations incorporated was 4.5). In addition, 3 random (non-oligonucleotide-derived) point mutations encoding an amino acid change were identified in 3636 codons (mutation rate of 0.08%). Approximately 5000 clones derived from the semi-synthetic shuffled library were screened for production of CHC-B1 and CHC-B2. Approximately 10% of colonies inoculated on the production plate did not grow, and approximately 20% of the analyzed clones did not produce CHC-B1 yields above that of the *aveC* deletion mutant. These percentages are consistent with those previously reported for this assay (Stutzman-Engwall et al., 2003). Among the remaining 70%, 89 were identified as producing a CHC-B2:CHC-B1 ratio that was improved over the previously identified best clone (Fig. 2). A 70% live rate of active variants in the library corresponds to a 3.5-fold oversampling of the theoretical library size ($2^{10} = 1024$), resulting in approximately 97% coverage of unique variants (Patrick et al., 2003).

The 89 promising variants were evaluated in production tests (30 ml liquid media). Duplicate HPLC testing determined that 90% of these variants produced a CHC-B2:CHC-B1 ratio in liquid growth that was consistent with the first tier assay (96 well plate using solid media), validating the miniaturized assay format and demonstrating the correlation between solid and liquid growth screen, and a correlation between detection by MS and

HPLC. Among these, eight variants reproducibly produced a CHC-B2:CHC-B1 at a ratio of 0.2:1.

DNA sequencing revealed that each of the eight improved second round variants were unique, encoding on average seven amino acid changes (Fig. 3, Table 1). The success of DNA shuffling relies on rapid fixation and accumulation of cumulative beneficial mutations from multiple parents, as well as the removal of deleterious mutations from the population (F99S, G111V). While the standard shuffling format is strongly biased toward the incorporation of fewer, potentially beneficial mutations, with the most likely number of mutations at 1 per gene, the semi-synthetic format demonstrates a more uniform probability distribution, with the most likely number of mutations at 5 per gene. In particular, the probability of incorporating 7 mutations per gene as seen here is over 10,000 times more likely when using the semi-synthetic approach, as compared to standard DNA shuffling. Furthermore, semi-synthetic shuffling will allow for uncoupling of mutations that are positioned close in sequence space and are unlikely to be recombined by traditional crossover events. In practice, a random mutation rate on the order of ~0.1% per base pair was observed in addition to changes introduced by oligonucleotides (this includes silent codon changes having no influence on the amino acid level).

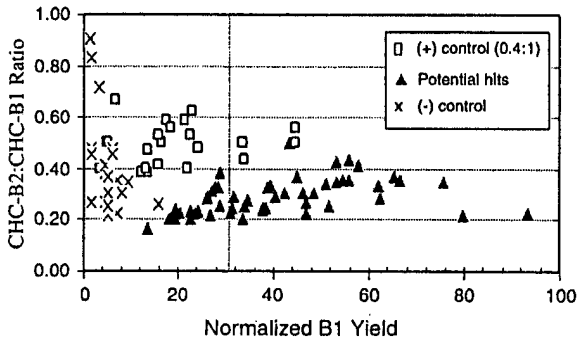


Fig. 2. Data points taken from the assay of a 96-well plate containing improved ratio variants after semi-synthetic DNA shuffling. The result is displayed as the ratio of CHC-B2:CHC-B1 as a function of the yield of CHC-B1. Clones were identified that produced CHC-B2:CHC-B1 ratios better than 0.4:1. The cut-off value for active clones was determined based on the highest yield CHC-B1 achieved from the negative control *aveC* deletion strain averaged over several experiments. Only clones producing > 30 μg/ml normalized B1 (vertical line) were analyzed further (normalized data derived from μg/ml production levels). The positive control is *S. avermitilis* strain SE180-11 transformed with pSE290 (D48E, A89T). The negative control is *S. avermitilis* strain transformed with the vector pWHM3.

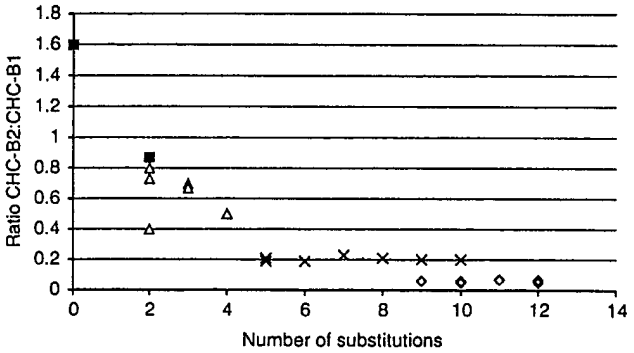


Fig. 3. CHC-B2:CHC-B1 ratio as a function of the number of substitutions per clone as compared to the wildtype *aveC* gene. Squares represent the two starting points; wildtype *aveC* and the S138T, A139T variant, pSE231. Open triangles represent the 7 best hits from the first round, X represent the 8 best hits from the second round, and open diamonds represent the 9 best hits from the third round.

3.2. Generation of additional genetic diversity

Semi-synthetic gene shuffling yielded a significant improvement in the CHC-B2:CHC-B1 ratio over clone pSE290 (0.4:1 to 0.2:1). Sequencing of 75 independent *aveC* clones from the semi-synthetic DNA shuffling reaction determined that at least two mutations were conserved in all cases (D48E and G179S, Table 1). It was not likely that the mutations introduced by semi-synthetic shuffling represented all possible beneficial genetic diversity that could be obtained in the *aveC* gene, so additional diversity was realized by constructing and screening additional libraries of randomly mutated *aveC* clones. This was accomplished by repeating error prone amplification of wt *aveC* in the presence of 0.2 mM manganese (Stutzman-Engwall et al., 2003) or Mutazyme. Approximately 3000 clones derived from these libraries were screened for production of CHC-B1 and CHC-B2; of these, 50 variants produced a CHC-B2:CHC-B1 ratio better than 1.3:1.

These 50 gene variants were evaluated in production tests (30 ml liquid media) and tested in duplicate by HPLC. Of these variants, 42 produced detectable CHC-B1 and 26 produced a CHC-B2:CHC-B1 ratio in liquid growth that was at least 1.3:1 or better (Fig. 4). DNA sequence analysis of the best 23 *aveC* clones determined that significant new genetic diversity was present. Several clones possessed a single mutation in *aveC* that resulted in an improved CHC-B2:CHC-B1 ratio of ~1:1. Ten new amino acid substitutions were identified that conferred improved CHC-B2:CHC-B1 ratios (S41G, R71L, L87V, W110L, L136M, T149S, F176C, L206M, G209R, and I280V). Oligonucleotides encoding these substitutions were designed so that all combinations of amino acid changes would be provided with equal probability. Oligonucleotides were combined and shuffled together with fragmented *aveC* derived from the best second round variant, pSE554 (Table 1).

3.3. Second and third rounds of semi-synthetic gene shuffling

Approximately 5000 clones from the second round of semi-synthetic shuffling were screened for production of

Table 1
Amino acid sequence of the best eight *aveC* variants generated after the first round of semi-synthetic DNA shuffling. The parent *aveC* gene used in the shuffling reaction was pSE290 and contained the mutations D48E and A89T. CHC-B2:CHC-B1 ratios are pSE506 0.19:1; pSE549 0.17:1; pSE554 0.18:1, pSE537 0.19:1; pSE505 0.2:1; pSE502 0.2:1; pSE503 0.19:1; pSE515 0.19:1

```
       1                                                                                                  100
 pSE506  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE549  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE554  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE537  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE505  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE502  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE503  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE515  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFTSWQSPLMNWFH
aveC WT  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTDVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFASWQSPLMNWFH

       101                                                                                                200
 pSE506  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLATFSVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
 pSE549  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLATTSVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
 pSE554  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLATTSVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
 pSE537  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLATTSVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
 pSE505  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLATTSVCMSALIVTVLCSEALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
 pSE502  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISVWARA
 pSE503  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVTVLCSEALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISVWARA
 pSE515  SVLVSNASVWGAVGSWGPYAPGWQGAGPGAEAEMPPASASVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISVWARA
aveC WT  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLASASVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFIGIVLGLSEPLPSASGISVWARA

       201                                                                                          303
 pSE506  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
 pSE549  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
 pSE554  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
 pSE537  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDHDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
 pSE505  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDERDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
 pSE502  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
 pSE503  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
 pSE515  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDERDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
aveC WT  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDERDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
```

CHC-B1 and CHC-B2; 67 variants produced a CHC-B2:CHC-B1 ratio better than 0.2:1. These improved variants were progressed through shake flask validation in liquid medium and tested in duplicate, as described above; 88% of these variants produced a CHC-B2:CHC-B1 ratio consistent with the first tier solid media assay. Of these, six reproducibly produced CHC-B2:CHC-B1 at a ratio of 0.1:1 (Fig. 5, Table 2). DNA sequence analysis determined that the clones contained a unique complement of mutations, although all six clones had acquired the mutations D48E, A89T, T149S, G179S, and E238D.

Further analysis of the variants from the random mutation library determined that three new amino acid substitutions conferred improved CHC-B2:CHC-B1 ratios (V2M, S90N, C142Y). Oligonucleotides encoding these substitutions, as well as the oligonucleotides from the second round of semi-synthetic shuffling, were designed so that all combinations of amino acid changes could be allowed with equal probability. The oligonucleotides were combined and shuffled with fragmented *aveC* derived from pSE554. Approximately 5000 clones from the third round of semi-synthetic shuffling were screened for production of CHC-B1 and CHC-B2; 96 were identified as producing a CHC-B2:CHC-B1 ratio better than 0.1:1.

The 96 improved variants were retested in liquid media, as described before. Eighty percent of the variants produced a CHC-B2:CHC-B1 ratio in liquid medium that was consistent with the solid media assay results. DNA sequence analysis was completed on the best 16 variants, all of which reproducibly produced CHC-B2:CHC-B1 at a ratio of 0.07:1 or less (Fig. 5, Table 3). Although each sequenced clone contained a unique complement of mutations, all sixteen clones had acquired the common mutations D48E, R71L, A89T, T149S, F176C, G179S, and E238D.

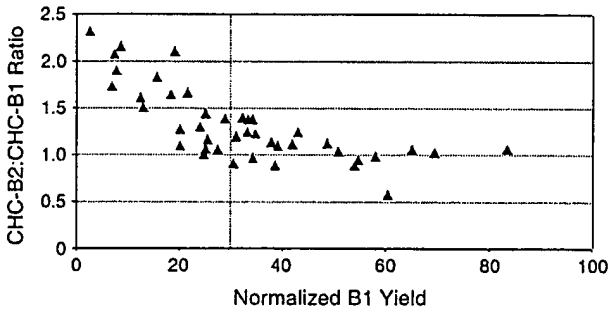


Fig. 4. Data points taken after fermentation analysis of *aveC* variants at 30 ml scale after a new round of random mutagenesis to identify new genetic diversity. The result is displayed as the ratio of CHC-B2:CHC-B1 as a function of the yield of CHC-B1 (normalized data derived from μg/ml production levels). The best 23 clones (CHC-B2:CHC-B1 ratios better than 1.3:1) were sequenced.

3.4. *Insertion of the CHC-B2:CHC-B1 0.07:1 mutation into the S. avermitilis production strain chromosome*

Two different approaches for gene replacement vector construction were chosen. The first approach used the *aveC* internal restriction fragment *PstI*/*SphI* (~660 bp) from the mutated *aveC* clone pSE617 to subclone into the gene replacement vector, as previously described (Stutzman-Engwall et al., 2003). The simplicity of this approach is attractive; however, mutations outside of the *SphI*/*PstI* fragment are not retained (see Table 3). The gene replacement plasmid pSE370 (containing the *aveC* gene with mutations D48E, A61T, R71L, A89T, L136P, T149S, F176C, G179S, and V196A) was transformed into protoplasts of *S. avermitilis* SE247-11 (a high titer *aveC* knockout strain) and analyzed by fermentation for avermectin production. Fermentation analysis showed that the pSE370 transformants were producing very low avermectin titers. A second gene replacement strain was constructed to introduce the mutation E238D (which was 3′ to the *SphI* site and not included in the subcloning of pSE370). The gene replacement plasmid pSE375 (containing the *aveC* gene variant with mutations D48E, A61T, R71L, A89T, L136P, T149S, F176C, G179S, V196A, and E238D) was

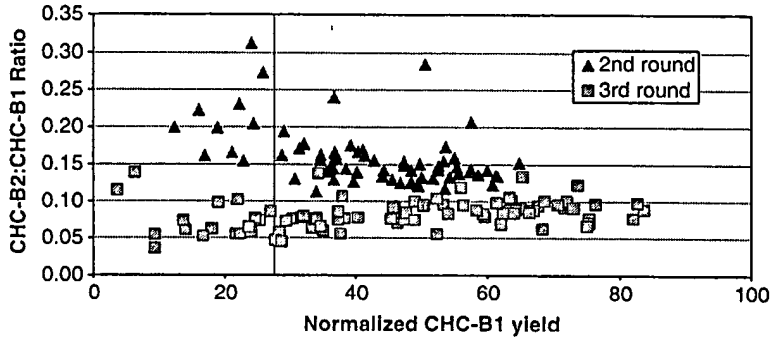


Fig. 5. Data points taken after fermentation analysis of improved *aveC* variants at 30 ml scale after the second and third rounds of semi-synthetic gene shuffling. Clones were identified that produced CHC-B2:CHC-B1 ratios better than 0.2:1 for second round (triangles) and better than 0.1:1 for the third round (squares). The cut-off value for active clones (vertical line) was determined based on the highest yield CHC-B1 achieved from the negative control *aveC* deletion strain averaged over several experiments (normalized data derived from μg/ml production levels).

Table 2
Amino acid sequence of the six best *aveC* variants generated after second round of semi-synthetic DNA shuffling. The parent *aveC* gene used in the shuffling reaction was pSE554 and contained the mutations D48E, A61T, S138T, A139T, G179S, V196A, E238D, and P289L. CHC-B2:CHC-B1 ratios are pSE603 0.13:1; pSE601 0.12:1; pSE607 0.13:1, pSE609 0.13:1; pSE608 0.12:1; pSE610 0.13:1

```
          1                                                                                                  100
pSE603    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
pSE601    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGGKDTGTTEVVYPVISVVCITTAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
pSE607    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
pSE609    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGRGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCLVERRLLFDALLFLGLVFTSWQSPLMNWFH
pSE608    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
pSE610    VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAWLFRRCRVERRLLFDALLFLGLVFTSWQSPLMNWFH
aveC WT  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTDVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFASWQSPLMNWFH

          101                                                                                                200
pSE603    SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISVWARA
pSE601    SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPMATTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
pSE607    SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPMATTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
pSE609    SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPMATTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
pSE608    SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLASTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
pSE610    SVLVSNASVLGAVGSWGPYVPGWQGAGPGAEAEMPLATTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVFFISIVLGLSEPLPSASGISAWARA
aveC WT  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLASASVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFIGIVLGLSEPLPSASGISVWARA

          201                                                                                          303
pSE603    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
pSE601    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCLHILLSLVGGQLPDQLPDSFQAPAAY*
pSE607    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
pSE609    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
pSE608    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLGGGQLPDQLPDSFQAPAAY*
pSE610    LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
aveC WT  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDERDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
```

Table 3
Amino acid sequence of the nine best *aveC* variants generated after the third round of semi-synthetic DNA shuffling. The parent *aveC* gene used in the shuffling reaction was pSE554 and contained the mutations D48E, A61T, S138T, A139T, G179S, V196A, E238D, and P289L. CHC-B2:CHC-B1 ratios are pSE659 0.06:1; pSE657 0.06:1; pSE670 0.07:1, pSE622 0.07:1; pSE646 0.06:1; pSE639 0.07:1; pSE617 0.05:1; pSE682 0.05:1 and pSE621 0.06:1. *pSE617 was used for gene replacements

```
          1            PstI                                                                100
 pSE659   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE657   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE670   VMVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE622   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE646   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE639   VMVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE617*  VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE682   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
 pSE621   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGPGQGGSKDTGTTEVVYPVISVVCITTAAAAWLFRRCLVERRLLFDALLFLGLLFTSWQSPLMNWFH
aveC WT   VVVWAGVGLLFLALQAYVFSRWAADGGYRLIETAGQGQGGSKDTGTTDVVYPVISVVCITAAAAWLFRRCRVERRLLFDALLFLGLLFASWQSPLMNWFH

          101                                                                          200
 pSE659   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE657   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE670   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE622   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLASTSVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE646   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISVWARA
 pSE639   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE617*  SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE682   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKAHRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
 pSE621   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPPASASVCMSALIVSVLCSKALGWIKARRPAWRTWRLVLAVCFISIVLGLSEPLPSASGISAWARA
aveC WT   SVLVSNASVWGAVGSWGPYVPGWQGAGPGAEAEMPLASASVCMSALIVTVLCSKALGWIKARRPAWRTWRLVLAVFFIGIVLGLSEPLPSASGISVWARA

          201                SphI                                                      303
 pSE659   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
 pSE657   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE670   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE622   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE646   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQLPDQLPDSFQAPAAY*
 pSE639   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE617*  LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE682   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
 pSE621   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDDRDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHVLLSLVGGQPPDQLPDSFQAPAAY*
aveC WT   LPEVTLWSGEWYQFPVYQAVGSGLVCCMLGSLRFFRDERDESWVERGAWRLPQRAANWARFLAVVGGVNAVMFLYTCFHILLSLVGGQPPDQLPDSFQAPAAY*
```

transformed into protoplasts of *S. avermitilis* SE247-11 and analyzed by fermentation for avermectin production. Fermentation analysis showed that the pSE375 transformants were producing normal quantities of avermectins at a CHC-B2:CHC-B1 ratio of 0.07:1. To incorporate mutations conferring improved product ratio into the chromosome, single colony pSE375 transformants of *S. avermitilis* SE247-11 were screened for gene replacement by monitoring the loss of thiostrepton and erythromycin resistance (Stutzman-Engwall et al., 2003). Fermentation analysis showed that Erm^S $Thio^S$ transformants produced avermectins at a 0.07:1 CHC-B2:CHC-B1 ratio (Fig. 6). Gene replacement was verified by PCR amplification of the *aveC* gene and DNA sequence analysis. No ErmE, Thio or pWHM3 sequences were detected after PCR amplification using appropriate primers (data not shown), confirming the mutated *aveC* gene integrated by double cross-over.

4. Discussion

The ability to carry out molecular genetic manipulations on high titer production strains used in industrial fermentations is an extremely desirable but highly challenging objective. Here, we present an important milestone toward the goal of engineering the biosynthetic pathway of a natural product (avermectin) for pharmaceutical applications. By integrating DNA shuffling with HTP electrospray mass spectrometry, we developed a sensitive and accurate method to monitor polyketide production and to manipulate the polyketide synthetic pathway while maintaining overall avermectin expression in the production organism. This combination of technologies allows for the creation, modification and production of polyketides in ways that are inaccessible to traditional medicinal chemistry.

Directed evolution has been highly successful for proteins where simple and rapid HTP screens or selections are available. No such tool is generally available for polyketide screening. In addition, a major limitation for HTP screening of small molecule production from *Streptomyces* has been to achieve reproducible cell growth and production levels of independent variants of secondary metabolites. These challenges were overcome by improving the quality of library diversity, thereby reducing the number of variants that needed to be screened, and by establishing a robust 96 well solid phase growth and production screen. We describe a semi-synthetic shuffling format in which diversity is identified by sequencing and subsequently incorporated by oligonucleotides into the shuffling reaction. Semi-synthetic shuffling enables each combination of amino acid substitutions from a single gene shuffled library to occur at an equal probability. This strategy proved effective in generating high quality libraries, ultimately generating clones with a 23-fold improved activity over the wild-type gene, and containing an average of 10 amino acid substitutions. This was achieved after four rounds of directed evolution and screening of 28,000 independent clones in total.

While we succeeded in creating a new *S. avermitilis* strain having an improved CHC-B2:CHC-B1 ratio, the function of the AveC protein in the polyketide synthesis production pathway remains unclear. Although a functional or structural understanding of an underlying biological system can assist in its improvement, protein optimization through DNA shuffling does not depend on it. Instead, as in this example, we only required sequence-function relationship: the information that amino acid substitutions in *aveC* change the

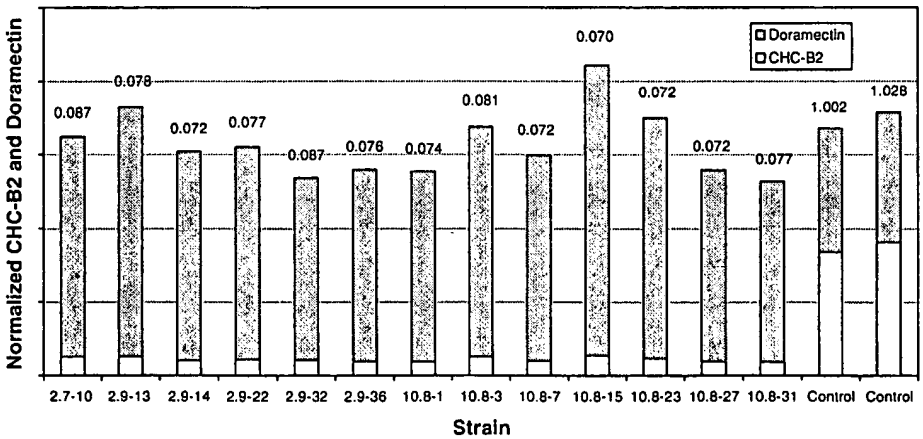


Fig. 6. Fermentation analysis of thirteen independent *S. avermitilis* transformants containing the mutated *aveC* allele (D48E, A61T, R71L, A89T, L136P, T149S, F176C, G179S, V196A, and E238D) introduced by gene replacement into the chromosome of the *S. avermitilis* production strain (normalized data derived from μg/ml production levels). The variant *aveC* gene directs the production of CHC-B2:CHC-B1 at a 0.07:1 ratio. The control strains produce CHC-B2:CHC-B1 at a 1:1 ratio.

CHC-B2:CHC-B1 ratio. By screening for an improved CHC-B2:CHC-B1 ratio and recombining the identified diversity, we have engineered a highly complex natural product pathway. The same approach can be used for other biological production pathways as long as a reproducible format for HTP growth and compound production is integrated with accurate analytical screens and DNA shuffling.

Acknowledgments

We thank Lu Liu, Jon Ness, Yong Hong Chen and Jeremy Minshull (Maxygen) Rick Davies and Jim Lengen (Pfizer) and Matt Tobin and Lori Giver (Codexis) for helpful discussions, expert technical assistance and critical reading of the manuscript.

References

Aita, T., Hamamatsu, N., Nomiya, Y., Uchiyama, H., Husimi, Y., 2002. Surveying a local fitness landscape of a protein with epistatic sites for the study of directed evolution. Biopolymers 64, 95–105.

Burg, R.W., Miller, B.M., Baker, E.E., Birnbaum, J., Currie, S.A., Hartman, R., Kong, Y.L., Monaghan, R.L., Olson, G., Putter, I., Tunac, J.B., Wallick, H., Stapley, E.O., Oiwa, R., Omura, S., 1979. Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. Antimicrob. Agents Chemother. 15, 361–367.

Cane, D.E., Liang, T.C., Kaplan, L.K., Nallin, M.K., Schulman, M.D., Hensens, O.D., Douglas, A.W., Albers-Schonberg, G., 1983. Biosynthetic origin of the carbon skeleton and oxygen atoms of the avermectins. J. Am. Chem. Soc. 105, 4110–4112.

Chartrain, M., Salmon, P.M., Robinson, D.K., Buckland, B.C., 2000. Metabolic engineering and directed evolution for the production of pharmaceuticals. Curr. Opin. Biotechnol. 11, 209–214.

Chen, T.S., Arison, B.H., Gullo, V.P., Inamine, E.S., 1989. Further studies on the biosynthesis of the avermectins. J. Ind. Microbiol. 4, 231–238.

Chen, T.S., Inamine, E.S., 1989. Studies on the biosynthesis of avermectins. Arch. Biochem. Biophys. 270, 521–525.

Crameri, A., Stemmer, W.P.C., Minshull, J., Bass, S.H., Welch, M., Ness, J.E., Gustafsson, C., Patten, P.A., 2000. Oligonucleotide mediated nucleic acid recombination. PCT International Patent Application WO 00/42561.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Dutton, C.J., Gibson, S.P., Goudie, A.C., Holdom, K.S., Pacey, M.C., Ruddock, J.C., Bu'Lock, J.D., Richards, M.S., 1991. Novel avermectins produced by mutational biosynthesis. J. Antibiot. 44, 357–365.

Goudie, A.C., Evans, N.A., Gration, K.A.F., Bishop, B.F., Gibson, S.P., Holdom, K.S., Kaye, B., Wicks, S.R., Lewis, D., Weatherley, A.J., Bruce, C.I., Herbert, A., Seymour, D.J., 1993. Doramectin—a potent novel endectocide. Vet. Parasitol. 49, 5–15.

Hafner, E.W., Holley, B.W., Holdom, K.S., Lee, S.E., Wax, R.G., Beck, D., McArthur, H.A.I., Wernau, W.C., 1991. Branched-chain fatty acid requirement for avermectin production by a mutant of *Streptomyces avermitilis* lacking branched-chain 2-oxo acid dehydrogenase activity. J. Antibiot. 44, 349–356.

Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., Omura, S., 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. USA 96, 9509–9514.

Ikeda, H., Omura, S., 1995. Control of avermectin biosynthesis in *Streptomyces avermitilis* for the selective production of a useful component. J. Antibiot. 48, 549–562.

Ikeda, H., Pang, C.-H., Endo, H., Ohta, T., Tanaka, H., Omura, S., 1995. Construction of a single component producer from the wild type avermectin producer *Streptomyces avermitilis*. J. Antibiot. 48, 532–534.

Katz, L., Donadio, S., 1993. Polyketide synthesis: prospects for hybrid antibiotics. Ann. Rev. Microbiol. 47, 875–912.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A., 2000. Practical Streptomyces Genetics. John Innes Foundation, Norwich.

MacNeil, D.J., Occi, J.L., Gewain, K.M., MacNeil, T., Gibbons, P.H., Ruby, C.L., Danis, S.J., 1992. Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. Gene 115, 119–125.

Moore, C.J., Jin, H-M., Kuchner, O., Arnold, F.A., 1997. Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol. 272, 336–347.

Patrick, W.M., Firth, A.E., Blackburn, J.M., 2003. User-friendly algorithms for estimating the completeness and diversity in randomized protein-encoding libraries. Protein Eng. 16, 451–457.

Pacey, M.S., Dutton, C.J., Monday, R.A., Ruddock, J.C., Smith, G.C., 2000. Preparation of 13-*epi*-Selamectin by biotransformation using a blocked mutant of *Streptomyces avermitilis*. J. Antibiot. 53, 301–305.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Stemmer, W.P.C., 1994a. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389–391.

Stemmer, W.P.S., 1994b. DNA in vitro shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA 91, 10747–10751.

Stutzman-Engwall, K.J., Conlon, S.W., Fedechko, R.W., Kaczmarek, F.S., 1997. Engineering *Streptomyces avermitilis* biosynthetic genes. In: Hutchinson, C.R., McAlpine, J. (Eds.), Developments in Industrial Microbiology Series, vol. 35. Society for Industrial Microbiology, Fairfax, VA, pp. 7–13.

Stutzman-Engwall, K.J., Conlon, S., Fedechko, R., Kaczmarek, F., McArthur, H., Krebber, A., Chen, Y., Minshull, J., Raillard, S.A., Gustafsson, C., 2003. Engineering the *aveC* gene to enhance the ratio of doramectin to its CHC-B2 analogue produced in *Streptomyces avermitilis*. Biotech. Bioeng. 82, 359–369.

Stutzman-Engwall, K.J., McArthur, H., Katoh, Y., 2001. *Streptomyces avermitilis* gene directing the ratio of B2:B1 avermectins. US Patent 6,248,579.

Vajdos, F.F., Adams, C.W., Breece, T.N., Presta, L.G., de Vos, A.M., Sidhu, S.S., 2002. Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. J. Mol. Biol. 320 (2), 415–428.

Vara, J., Lewandowska-Skarbek, M., Wang, Y.-G., Donadio, S., Hutchinson, C.R., 1989. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthetic pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). J. Bacteriol. 171, 5872–5881.